(3.4)

In single variable calculus we learned how to find the maxima and minima of curves lying on a flat, two-dimensional surface.

Fig. 1

In higher dimensional settings however, this surface might meander through space:

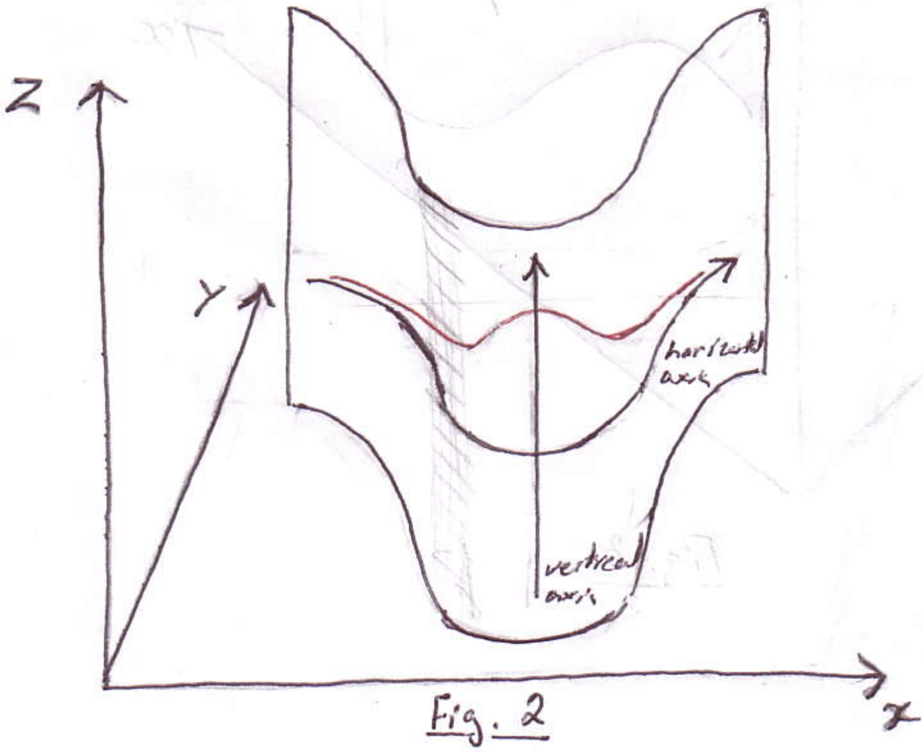


Fig. 2

Here, the curve on the contorted surface is some function of the "horizontal axis". That is, when flattened, the curve will look like the picture above (Fig. 1).
Notice that the surface is a cylinder projected along the z-axis. The curve may therefore be viewed as the intersection of the cylinder surface with some function $z = f(x,y)$.

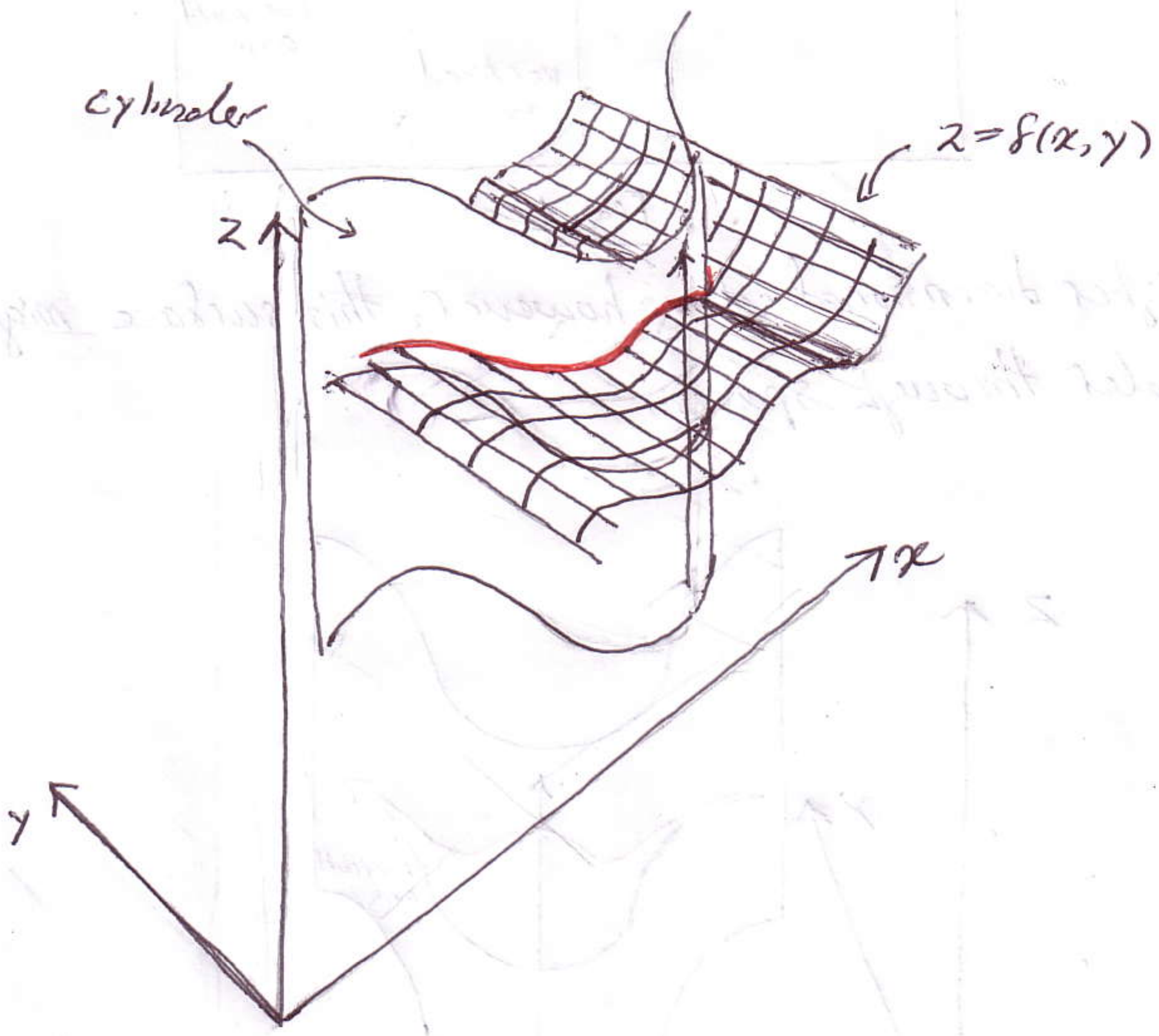


Fig. 3

Observe that we can describe the cylinder surface by some equation of the form $g(x,y)=0$. For example, if the surface is a unit circular cylinder then it is identified by the equation $x^2+y^2-1=0$ where $g(x,y)=x^2+y^2-1$, whereas if the cylinder is generated by projecting the curve $y=\sin x+3\cos x$ along the z-axis, the surface is identified by the equation $y-\sin x-3\cos x=0$ and we set $g(x,y)=y-\sin x-3\cos x$. With this notation, we can determine the red curve in Fig. 3 as the intersection of the cylinder surface $G_g=\{(x,y,z): g(x,y)=0\}$ with the set $G_f=$

$\{(x,y,z): z=f(x,y)\} = \{(x,y,f(x,y)): (x,y)\in\mathbb{R}^2\}$.

In other words, the curve is the set $C=G_g\cap G_f$ or

$C=\{(x,y,f(x,y)): g(x,y)=0\}$. Geometrically, the curve is just the ordinary graph of f restricted to a domain D that is implicitly defined by the restriction $g(x,y)=0$.

The problem of finding the maxima and minima of the curve C can thus be stated as follows:

Find the maxima and minima of the function $z=f(x,y)$ under the constraint $g(x,y)=0$.

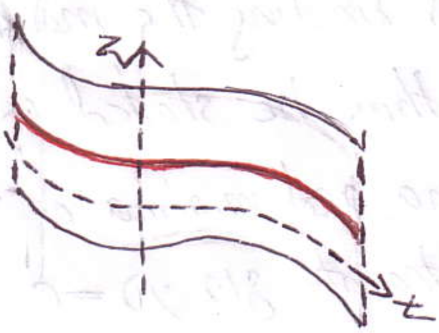
The mathematics that deals with this problem is known as the field of constrained optimization.

Ex. Indentify the maxima and minema of the function $f(x,y) = (y-x^3)^2$ under the constraint $y - x^3 = 2$.

Solution: let $D = \{(x,y) \in \mathbb{R}^2 : y - x^3 - 2 = 0\}$

We wish to find the maxima and minema of the function $f: D \to \mathbb{R}$. This is very easy, because f is constant on D: for any point $(x,y) \in D$ $y - x^3 = 2$ so $f(x,y) = (y-x^3)^2 = 2^2 = 4$. Thus, every point in D is both a maximum and a minimum.

This conclusion may also be derived geometrically: Recall that the graph of $f(x,y) = (y-x^3)^2$ is a cylinder-like structure that is obtained by projecting the curve $z = y^2$ along the "axis" $y = x^3$. It follows that the function is constant along any curve that is parallel to $y = x^3$. The graph of $f: D \to R$ is just the line $z = 4$ on the coiled plane zt's where t is the axis $y = x^3 + 2$



Ex. Find the maxima and minima of the function $f(x,y)=xy$ under the constraint $x^2+\frac{y^2}{9}=1$.

Solution: We wish to optimize the function $f: D\subset \mathbb{R}^2 \to \mathbb{R}$ where $D=\{(x,y): x^2+\frac{y^2}{9}=1\}$.

We can parametrize D with the path function $c:[0,2\pi]\to\mathbb{R}^2$ given by $c(t)=(\cos t, 3\sin t)$. In other words, for each point $(x,y)\in D$, there exists some $t\in[0,2\pi]$ such that $(x,y)=(\cos t, 3\sin t)$.

Consider the function $p:[0,2\pi]\to\mathbb{R}$ given by $p(t)=f(c(t))$. Intuitively, you may think of the graph of p as the uncoiled graph of $f: D\subset\mathbb{R}^2\to\mathbb{R}$ (Just like Fig 1 is the uncoiled graph of Fig. 2). Therefore, if t_e is a point over which p has a local extremum, f will have a local extremum over the point $(x_e, y_e)=(\cos t_e, 3\sin t_e)$.

Now $p(t)=f(\cos t, 3\sin t)=3\cos t\sin t=\frac{3}{2}\sin 2t$.

Observe that $-\frac{3}{2}\leq\frac{3}{2}\sin 2t\leq\frac{3}{2}$. Thus, p attains its maximum value when $p(t)=\frac{3}{2}$. This happens when $t=\frac{\pi}{4}$ or when $t=\frac{\pi}{4}+\pi=\frac{5\pi}{4}$. Similar reasoning leads to the conclusion that p attains its minimum value of $-\frac{3}{2}$ when $t=\frac{3\pi}{4}$ or when $t=\frac{7\pi}{4}$.

Thus, $f: D\to\mathbb{R}$ attains its maximum at the points

$$(x,y) = \left(\cos\frac{\pi}{4}, 3\sin\frac{\pi}{4}\right) = \left(\frac{\sqrt{2}}{2}, 3\frac{\sqrt{2}}{2}\right);$$

$$(x,y) = \left(\cos\frac{5\pi}{4}, 3\sin\frac{5\pi}{4}\right) = \left(-\frac{\sqrt{2}}{2}, -3\frac{\sqrt{2}}{2}\right)$$

Similarly, the minimum value is reached at

$$(x,y) = \left(-\frac{\sqrt{2}}{2}, 3\frac{\sqrt{2}}{2}\right);$$

$$(x,y) = \left(\frac{\sqrt{2}}{2}, -3\frac{\sqrt{2}}{2}\right)$$

Unfortunately, a parametric description of the constraint $g(x,y)=0$ may lead to unwieldy formulas or worse still, we may not have the means to produce such description. For example, suppose that we wish to parametrize the constraint $y^3+x^2y+x^6=0$. With help from Cardan's cubic formula, we can solve for y to obtain

$$y(x) = \sqrt[3]{\frac{1}{2}\left(-x^6+\sqrt{x^{12}+\frac{4x^6}{27}}\right)} + \sqrt[3]{\frac{1}{2}\left(-x^6-\sqrt{x^{12}+\frac{4x^6}{27}}\right)}$$

The constraint may thus be parametrized by

$$c(t) = \left(t, \sqrt[3]{\frac{1}{2}\left(-t^6+\sqrt{t^{12}+\frac{4t^6}{27}}\right)} + \sqrt[3]{\frac{1}{2}\left(-t^6-\sqrt{t^{12}+\frac{4t^6}{27}}\right)}\right)$$

And we are dismayed by the monster that we have created. As ugly as the above function is, we wouldn't have advanced even to this stage if the constraint were $y^5+x^2y+x^6=0$, although, on first glance, it is barely distinguishable from its predecessor.

The method of Lagrange multipliers, wich we present below, offers an alternative approach to the problem of constrained optimization. It should be considered whenever the explicit method of parametrizing the constraint and thereby reducing the problem to single-variable calculus is inconvenient.

Let $f(x,y)$ be a scalar-valued function that we wish to optimize under the constraint $g(x,y)=0$.

Let $D=\{(x,y): g(x,y)=0\}$ and suppose that $a=(a_1,a_2)\in D$ is a point at which the restricted function $f: D\to\mathbb{R}$ has a potential maximum or minimum. That is, suppose that for some (possibly unknown) parametrization $c: I\to D$ near $a\in D$, the function $p(t)=f(c(t))$ has the derivative $p'(t)=0$ at $t=t_a$, where $c(t_a)=a$. By the chain rule

$$p'(t)=\nabla f(c(t))\cdot c'(t)$$

so

$$p'(t_a)=0$$

implies that $\nabla f(c(t_a))\cdot c'(t_a) = \nabla f(a)\cdot c'(t_a)=0$.

If $c'(t_a)\neq 0$, then geometrically the above statement implies that the gradient of f at a is orthogonal to the tangent line to the constraint at $(x,y) = (a_1, a_2)$ (See fig. 4).

Observe now that $c(I) \subset D$. This means that the range of c is lying in $D = \{(x,y) : g(x,y) = 0\}$. Therefore, $g(c(t)) = 0$ for every $t \in I$. By the chain rule,

$$\frac{d}{dt}\left[g(c(t))\right] = \nabla g(c(t)) \cdot c'(t) = 0$$

for every $t \in I$.

In particular, $\nabla g(c(t_a)) \cdot c'(t_a) = \nabla g(a) \cdot c'(t_a) = 0$. Again, if $c'(t_a) \neq 0$, this means that $\nabla g(a)$ is orthogonal to $c'(t_a)$. But if

$$\nabla f(a) \perp c'(t_a)$$

and

$$\nabla g(a) \perp c'(t_a)$$

where $\nabla f(a)$ and $\nabla g(a)$ are vectors in $\mathbb{R}^2$, we must conclude that they are parallel vectors. Now, since parallel vectors are linearly dependent, there must be some scaler λ such that

$$\nabla f(a) = \lambda \nabla g(a).$$

This scalar λ is known as Lagrange multiplier. We summarize our work in the following theorem:

Thm (Lagrange multipliers in two variables and one constraint): Suppose that $f|_g : D \to \mathbb{R}$ is the restriction of some function $f: \mathbb{R}^2 \to \mathbb{R}$ by the constraint $g(x,y) = 0$. Then $a \in D$ is a critical point of $f|_g$ if $\nabla f(a)$ is undefined or if

$$\nabla f(a) = \lambda \nabla g(a)$$

for some scalar $\lambda \in \mathbb{R}$.

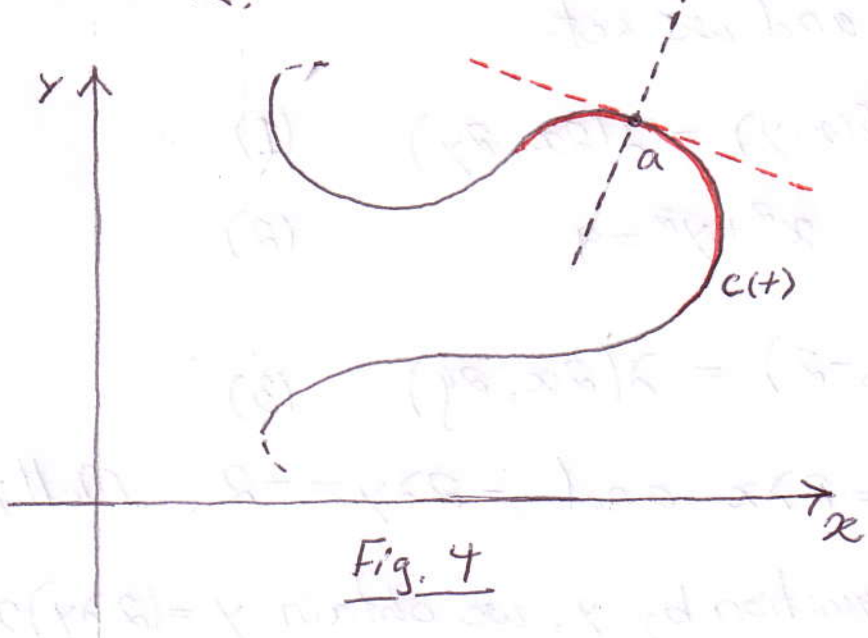


Fig. 4

— D (or $\{(x,y) : g(x,y) = 0\}$)

— $c(I) \subset D$

-- tangent at a generated by $c'(t_a)$

-- Line through a orthogonal to $c'(t_a)$

Fig. 4 illustrates the fact that in $\mathbb{R}^2$ each line through $a \in \mathbb{R}^2$ has a unique orthogonal line that intersects it at a. Thus, all vectors perpendicular to $c'(t_a)$ are linearly dependent.

Ex. Find the maxim and the minimum values of $f(x,y) = x - 2y$ subject to the constraint equation $x^2+y^2=9$.

Solution: Observe that f is a linear map. Its gradient is $\nabla f(x,y) = (1,-2)$. Let $g(x,y) = x^2+y^2-9$. Then the constraint equation is $g(x,y)=0$. Now $\nabla g(x,y) = = (2x, 2y)$ and we set

$$\nabla f(x,y) = \lambda(2x, 2y) \qquad (1)$$

$$x^2+y^2=9 \qquad (2)$$

$$(1,-2) = \lambda(2x, 2y) \qquad (3)$$

Thus, $1 = 2\lambda x$ and $2\lambda y = -2$. Multiplying the first equation by y, we obtain $y = (2\lambda y)x$.
By the second equation, $2\lambda y = -2$. Thus, we obtain the equation $y = -2x$. Substituting the latter equation into the constraint (eq. (2)), we get

$$x^2 + (-2x)^2 = 9 \quad \text{or}$$

$$x = \pm \frac{3}{\sqrt{5}}$$

Since $y=-2x$, we see that our critical points are $\left(\frac{-3}{\sqrt5}, \frac{6}{\sqrt5}\right)$ and $\left(\frac{3}{\sqrt5}, \frac{-6}{\sqrt5}\right)$

Notice that $f\left(\frac{-3}{\sqrt{5}}, \frac{6}{\sqrt{5}}\right) = \frac{1}{\sqrt{5}}(-3-2\cdot 6) = \frac{-15}{\sqrt{5}}$, while $f\left(\frac{3}{\sqrt{5}}, -\frac{6}{\sqrt{5}}\right) = \frac{15}{\sqrt{5}}$. Thus, f attains a maximum value of $\frac{15}{\sqrt{5}}$ over the point $\left(\frac{3}{\sqrt{5}}, -\frac{6}{\sqrt{5}}\right)$ and a minimum value of $-\frac{15}{\sqrt{5}}$ over the point $\left(\frac{-3}{\sqrt{5}}, \frac{6}{\sqrt{5}}\right)$.

Before we go on, several remarks are in order.

Remark 1

The problem we just solved could easily be dealt with by the direct method. The function we just optimized may be thought of as the real-valued single-variable function $p(t) = 3\cos t - 6\sin t$, because the constraint equation $x^2+y^2=9$ represents the set $\{(x,y,z) : x^2+y^2=9\} = \{(3\cos t, 3\sin t, z) : t \in [0, 2\pi]\}$

Remark 2

When you are solving a problem with the method of Lagrange multipliers, be sure to keep an inventory of your old equations. A common mistake is to dump the older equation, once it is used in the derivation of a newer, sexier one.

Remark 3

The method of Lagrange multipliers is designed to identify critical points. These critical points are not necessarily extreme points. That is, if $\nabla f(x_0, y_0) = \lambda g(x_0, y_0)$

for some point (x_0, y_0) that satisfies $g(x_0, y_0)=0$, it doesn't mean that $f|_g$ has a maximum or a minimum over (x_0, y_0).

Ex. Find the maximum and minimum of $f(x,y) = \sin(x^2y)$ subject to the constraint $y^3+x^2y+1=0$.

Solution: Let $g(x,y) = y^3+x^2y+1$. Then

$$\nabla g(x,y) = (2xy,\ 3y^2+x^2)$$

$$\nabla f(x,y) = (2xy\cos(x^2y),\ x^2\cos(x^2y)).$$

With the Lagrange multiplier condition, we obtain the following equations.

$$2xy\cos(x^2y) = \lambda(2xy) \qquad (1)$$

$$x^2\cos(x^2y) = \lambda(3y^2+x^2) \qquad (2)$$

$$y^3+x^2y+1=0 \qquad (3)$$

Observe that equation (3) implies that $y \neq 0$. Notice also that if $x=0$ then $y=-1$ by equation (3). At $(0,-1)$, $f(0,-1)=0$. Assume that $x \neq 0$. Then equation (1) reduces to

$$\cos(x^2y) = \lambda \qquad (4)$$

Plugging this value into (2), we obtain

$$x^2\cos(x^2y) = \cos(x^2y)(3y^2+x^2) \quad (5)$$

If $\cos(x^2y) \neq 0$, equation (5) becomes

$$x^2 = 3y^2+x^2 \quad \text{or} \quad y=0$$

Since y cannot be equal to 0, we must assume that $\cos(x^2y)=0$.

Thus $x^2y=\frac{\pi}{2}+2\pi n$, for $n\in\mathbb{Z}$ or $x^2y=\frac{3\pi}{2}+2\pi n$, for $n\in\mathbb{Z}$. (6).

Notice that we can rewrite equations (6) using equation (3): $y^3+x^2y+1=0$ implies that $x^2y=-(y^3+1)$.

Substituting the latter into equations (6), we obtain

$$-(y^3+1)=\frac{\pi}{2}+2\pi n \quad \text{or}$$

$$-(y^3+1)=\frac{3\pi}{2}+2\pi n.$$

Hence $y=-\left(\frac{\pi}{2}+2\pi n+1\right)^{1/3}$ or $y=-\left(\frac{3\pi}{2}+2\pi n+1\right)^{1/3}$. Since $x^2y=\frac{3\pi}{2}+2\pi n$, it follows that $x^2=-\left(\frac{3\pi}{2}+2\pi n\right)\left(\frac{3\pi}{2}+2\pi n+1\right)^{-1/3}$.

However, there is a problem: If $n<0$, $-\left(\frac{3\pi}{2}+2\pi n\right)>0$ and $\left(\frac{3\pi}{2}+2\pi n+1\right)^{-1/3}<0$, implying that $x^2<0$, which is impossible. If $n\geq 0$, $-\left(\frac{3\pi}{2}+2\pi n\right)<0$, while $\left(\frac{3\pi}{2}+2\pi n+1\right)^{-1/3}>0$, again implying that $x^2<0$. A similar analysis shows that the solution $y=-\left(\frac{\pi}{2}+2\pi n+1\right)^{1/3}$ does not work in conjunction with the equation $x^2y=\frac{\pi}{2}+2\pi n$.

After all this work, we are rewarded with the only critical point is $(0,-1)$. Is this point a local max or min? Notice that for points (x,y) satisfying the constraint $y^3+x^2y+1=0$, $\sin(x^2y)=\sin(-[y^3+1])=-\sin(y^3+1)$. If (x,y) is close to the point $(0,-1)$, then, in particular, y is close to -1 and y^3+1 is close to 0. If $y>-1$, $-\sin(y^3+1)<0$ while if $y<-1$, $-\sin(y^3+1)>0$. Thus $(0,-1)$ is a saddle point. It is neither a maximum nor a minimum.

Our lengthy analysis reveals that $\sin(x^2y)$ constrained to $y^3+x^2y+1=0$ has neither local max nor local min. The price was too high for the journey!

How can we tell when a constrained function $f|_g$ has extreme points? If the curve $g(x,y)=0$ is closed (for instance, a circle), then the maximum and minimum values of $f|_g$ occur among the points that are solutions to

$$\nabla f(x,y) = \lambda \nabla g(x,y)$$
$$g(x,y)=0.$$

In other words, if $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_m$ are the points that satisfy these equations, then

$$\max_{g(x,y)=0} f(x,y) = \max\{f(\vec{x}_1), f(\vec{x}_2), \ldots, f(\vec{x}_m)\}$$

and

$$\min_{g(x,y)=0} f(x,y) = \min\{f(\vec{x}_1), f(\vec{x}_2), \ldots, f(\vec{x}_m)\}$$

More generally, if the curve $g(x,y)=0$ is bounded and its complement is an open set, then we are assured of the existance of extreme points by the exteme-value theorem. The proof of this theorem requires the topological concept of compactness, which you will study in analysis. For now, it is beyond our scope.

There is one more issue that we should address before we can generalize Lagrange's strategy for identifying critical points. Suppose $\vec{a} = (a,b)$ is a solution to

$$\nabla f(x,y) = \lambda \nabla g(x,y)$$
$$g(x,y) = 0$$

corresponding to a scalar λ_0. That is, suppose

$$\nabla f(a,b) = \lambda_0 \nabla g(a,b)$$
$$g(a,b) = 0$$

we would like to develop a test to determine whether (a,b) is a local max, min or neither.

Thm. Let $\vec{a}$ and λ_0 be solutions to

$$\nabla f(x) = \lambda \nabla g(x)$$
$$g(x) = 0$$

and let $W(x,\lambda) = Hf(x) - \lambda Hg(x)$, where Hf and Hg are the Hessian matrices of the respective functions.

1. If $W(\vec{a},\lambda_0)$ is positive definite, then $f(\vec{a})$ is a local minimum value of f along $g(x,y)=0$.

2. If $W(\vec{a},\lambda_0)$ is negative definite, then $f(\vec{a})$ is a local maximum value of f along $g(x,y)=0$

proof: Let $x = c(t)$ be a parametrization of the level curve $g(x) = 0$ such that $c(t_a) = \vec{a}$ and $c'(t_a) \neq 0$. Let $p(t) = f(c(t))$. Then, by the second-derivative test of single-variable calculus, $f(\vec{a}) = p(t_a)$ is a local minimum value if $p'(t_a) = 0$ and $p''(t_a) > 0$, while $p(t_a)$ is a local maximum if $p''(t_a) < 0$.

Now,

$$p'(t) = \frac{d}{dt}\left[f(c(t))\right] = \nabla f(c(t)) \cdot c'(t).$$

so

$$p''(t) = \frac{d}{dt}\left[\nabla f(c(t)) \cdot c'(t)\right] = \frac{d}{dt}\left[\sum_{i=1}^{2} \frac{\partial f}{\partial x_i}(c(t))c_i'(t)\right] =$$

$$= \sum_{i=1}^{2} \frac{\partial f}{\partial x_i}(c(t))c_i''(t) + \left(\nabla \frac{\partial f}{\partial x_i}(c(t)) \cdot c'(t)\right)c_i'(t) =$$

$$= \sum_{i=1}^{2} \frac{\partial f}{\partial x_i}(c(t))c_i''(t) + \sum_{i=1}^{2}\left(\nabla \frac{\partial f}{\partial x_i}(c(t)) \cdot c'(t)\right)c_i'(t) =$$

$$= \nabla f(c(t)) \cdot c''(t) + c'(t)\, Hf(c(t))(c'(t))^T.$$

Therefore

$$p''(t_a) = \nabla f(\vec{a}) \cdot c''(t_a) + c'(t_a)\, Hf(\vec{a})\,(c'(t_a))^T.$$

Letting $G(t) = g(c(t))$, we see that $G(t) = 0$ (why?)

Notice that

$$G''(t_a) = \nabla g(\vec{a}) \cdot c''(t_a) + c'(t_a)\, Hg(\vec{a})\,(c'(t_a))^T$$ by

similar reasoning to the derivation of $p''(t_a)$.

Since $G(t)=0$ for all t, it follows that $G''(t_a)=0$. This latter equation can be rewritten as

$$\nabla g(\vec{a})\cdot c''(t_a) = -c'(t_a)Hg(\vec{a})(c'(t_a))^T \qquad (1)$$

We know, by hypothesis, that $\nabla f(\vec{a}) = \lambda_0 \nabla g(\vec{a})$. (2)

Therefore, upon multiplying equation (1) by λ_0 and making use of identity (2) we obtain

$$\nabla f(a)\cdot c''(t_a) = -c'(t_a)\lambda_0 Hg(\vec{a})(c'(t_a))^T.$$

Therefore,

$$\begin{aligned} p''(t_a) &= \nabla f(\vec{a})\cdot c''(t_a) + c'(t_a)Hf(\vec{a})(c'(t_a))^T = \\ &= -c'(t_a)\lambda_0 Hg(\vec{a})(c'(t_a))^T + c'(t_a)Hf(\vec{a})(c'(t_a))^T = \\ &= c'(t_a)\left(Hf(\vec{a}) - \lambda_0 Hg(\vec{a})\right)(c'(t_a))^T = \\ &= c'(t_a)\,W(\vec{a},\lambda_0)\,(c'(t_a))^T \end{aligned}$$

Since $c'(t_a)\neq 0$, we know from chapter 3.3 that $p''(t_a)>0$ whenever $W(\vec{a},\lambda_0)$ is positive-definite and $p''(t_a)<0$ whenever $W(\vec{a},\lambda_0)$ is negative-definite.

Ex. Find the maximum value of $f(x,y)=y-x^2$ subject to the constraint $x-y^2=0$

Solution: $\nabla f(x,y)=(-2x,1)$ and $\nabla g(x,y)=(1,-2y)$, so we must solve

$$\begin{aligned} -2x &= \lambda && (1)\\ 1 &= -2\lambda y && (2)\\ x-y^2 &= 0 && (3) \end{aligned}$$

By rewriting equation (3) into the form

$$x = y^2 \quad (4)$$

We replace x by y^2 in equation (1) to obtain

$$-2y^2 = \lambda \quad (5).$$

Plugging this identity into equation (2) we get

$$1 = -2(-2y^2)y = 4y^3 \quad (6)$$

Thus, $y = 4^{-1/3}$, $x = 4^{-2/3}$ and $\lambda_0 = -2 \cdot 4^{-2/3}$. In particular, our only critical point is $\vec{a} = (4^{-2/3}, 4^{-1/3})$.

The Hessian matrices are

$$Hf(x,y) = \begin{pmatrix} -2 & 0 \\ 0 & 0 \end{pmatrix} \text{ and } Hg(x,y) = \begin{pmatrix} 0 & 0 \\ 0 & -2 \end{pmatrix} \text{ so}$$

$$W\left((4^{-2/3}, 4^{-1/3}), -2 \cdot 4^{-2/3}\right) = Hf(4^{-2/3}, 4^{-1/3}) - (-2 \cdot 4^{-2/3}) Hg(4^{-2/3}, 4^{-1/3}) =$$

$$= \begin{pmatrix} -2 & 0 \\ 0 & -2 \cdot 4^{-2/3} \end{pmatrix}$$

This is negative definite, so the maximum of f subject to the constraint is $f(4^{-2/3}, 4^{-1/3})$

Thus far, we have examined the technique of Lagrange multipliers in the special case of two variables and one constraint. In the remainder of this chapter, we explore this technique in more general settings.

Three variables and one constraint

Suppose that we wish to optimize $f(x,y,z)$ under the constraint $g(x,y,z)=0$. The constraint equation implicitly defines a surface, which can be parametrized by a function of two independent variables. For example, suppose that we are given the constraint equation $x^2+y^2+z^2-4=0$. Then this equation defines a sphere of radius 2 with center at the origin. We can parametrize the sphere by the (surface) function $c: [0,2\pi]\times[0,\pi] \to \mathbb{R}^3$ given by $c(s,t) = (2\sin t \cos s, 2\sin t \sin s, 2\cos t)$.

When we are able to find an explicit parametrization $c(s,t)$ of the constraint surface $g(x,y,z)=0$, the task of locating the extreme values of $f|_g$ reduces to ordinary optimization of a scalar-valued function of two variables. Namely, optimizing $f|_g$ is equivalent to the task of finding the maxima and minima of $p(s,t) = f(c(s,t))$.

when the surface function $c(s,t)$ is not known or the composition $p(s,t) = f(c(s,t))$ is difficult to optimize, we may resort

to a different strategy:
Observe that $g(x,y,z)=0$ is a short-hand for $S_0 = \{(x,y,z): g(x,y,z)=0\}$, which is the 0-level surface of $g: \mathbb{R}^3 \to \mathbb{R}$. Recall from chapter 2.6 that the gradient of a differentiable function is perpendicular to the function's level surfaces. More precisely, if $\vec{a}=(a_1,a_2,a_3) \in S_0$, then $\nabla g(\vec{a})$ is perpendicular to the tangent plane to the surface S_0 at $\vec{a}$ (see Fig. 5).
Suppose now that $f|_g$ has a local max or a local min at $\vec{x}=\vec{a}$. Let $C(s,t)$ be a (possibly unknown) parametrization of S_0 with $C(s_a,t_a)=\vec{a}$ for some $(s_a,t_a)\in\mathbb{R}^2$. Define $u(s)=C(s,t_a)$ and $v(t)=C(s_a,t)$. Then u and v are path functions.
Let $k(s)=f(u(s))$ and $p(t)=f(v(t))$. Observe that $k(s_a)=p(t_a)=f(\vec{a})$ is an extreme value of $f|_g$. This means that $k(s_a)$ is an extreme value of the single-variable scalar function k and $p(t_a)$ is an extreme value of the single-variable scalar function p. In particular $\frac{dk}{ds}(s_a)=0$ and $\frac{dp}{dt}(t_a)=0$ (why?)
By the chain rule, this implies that

$$0=\frac{dk}{ds}(s_a)=\nabla f(\vec{a})\cdot u'(s_a) \qquad (1)$$

$$0=\frac{dp}{dt}(t_a)=\nabla f(\vec{a})\cdot v'(t_a) \qquad (2)$$

If $\|u'(s_a)\times v'(t_a)\|\neq 0$, we may conclude that

1. $u'(s_a)$ and $v'(t_a)$ are linearly independent vectors in $\mathbb{R}^3$. In particular, $u'(s_a) \neq 0$ and $v'(t_a) \neq 0$.

2. $\nabla f(\vec{a})$ is orthogonal to any vector of the form $s\,u'(s_a) + t\,v'(t_a)$, $s, t \in \mathbb{R}$. In particular, $\nabla f(\vec{a})$ is orthogonal to the tangent plane to the surface S_0 at $\vec{a}$.

3. Recall that $\nabla g(\vec{a})$ is also orthogonal to the tangent plane to the surface S_0 at $\vec{a}$. Since all vectors in $\mathbb{R}^3$ perpendicular to a plane must be linearly dependent, it follows that $\nabla f(\vec{a})$ and $\nabla g(\vec{a})$ are linearly dependent. Thus there must be a scalar $\lambda \in \mathbb{R}$ such that $\nabla f(\vec{a}) = \lambda \nabla g(\vec{a})$.

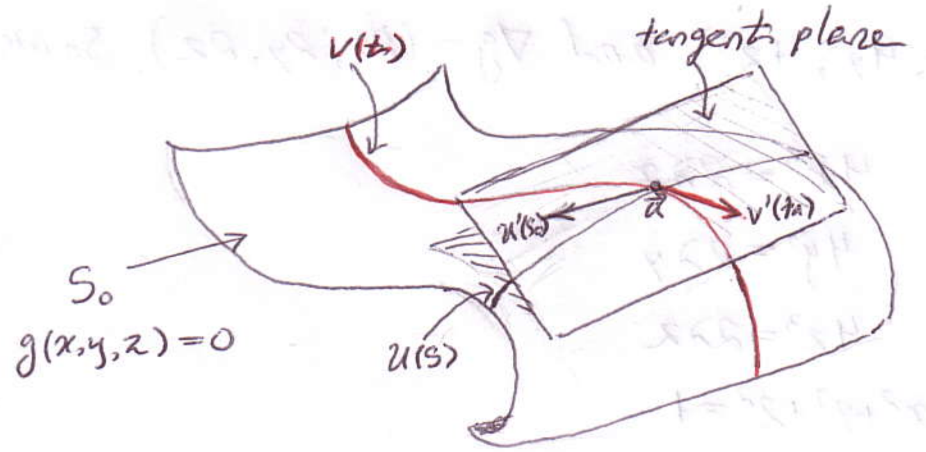


Fig. 5

We have proven the following theorem:

Thm (Lagrange multipliers in three variables and one constraint): Suppose that $f|_g : D \to \mathbb{R}$ is the restriction of some function $f: \mathbb{R}^3 \to \mathbb{R}$ by the constraint $g(x,y,z) = 0$. Then $a \in D$ is a critical point of $f|_g$ iff $\nabla f(a)$ is undefined or if

$$\nabla f(a) = \lambda \nabla g(a)$$

for some scaler $\lambda \in \mathbb{R}$.

Ex. (Thomas H. Barr). Find the minimum value of $f(x,y,z) = x^4+y^4+z^4$ on the sphere $x^2+y^2+z^2=1$.

Solution: The constraint set is closed and bounded, so f attains its minimum value on the sphere. We simply find all solutions of the system

$$\nabla f(x) = \lambda \nabla g(x)$$
$$g(x) = 0$$

and evaluate f at each one. The point(s) yielding the smallest value will be the locations we seek.

First, $\nabla f = (4x^3, 4y^3, 4z^3)$ and $\nabla g = (2x, 2y, 2z)$. So we must solve

$$4x^3 = 2\lambda x$$
$$4y^3 = 2\lambda y$$
$$4z^3 = 2\lambda z$$
$$x^2+y^2+z^2 = 1$$

Equivalently,

$$x(2x^2-\lambda) = 0$$
$$y(2y^2-\lambda) = 0$$
$$z(2z^2-\lambda) = 0$$
$$x^2+y^2+z^2 = 1$$

Not all of x, y, and z can be zero (why?), but having any pair of these variables equal zero gives a solution.

For instance, $x=0$ and $y=0$ satisfy the first two equations. These values in the fourth equation give $z=\pm 1$, so by the third equation $\lambda=2$. Thus $(0,0,\pm 1)$ with $\lambda=2$ are two points at which the minimum may occur. Similarly, $(0,\pm 1,0)$ with $\lambda=2$ and $(\pm 1,0,0)$ with $\lambda=2$ are four other points where the minimum may occur.

Second, consider the instances where exactly one of x, y, or z is zero. If $x=0$, and y and z are not zero, then by the second and third equations, $y^2=z^2$. Substituting this into the fourth equation, we have $2y^2=1$, so $y=\pm 1/\sqrt{2}$. So we obtain $(0, 1/\sqrt{2}, \pm 1/\sqrt{2})$ and $(0, -1/\sqrt{2}, \pm 1/\sqrt{2})$ as possible locations of minima. (What are the corresponding values of λ?). Similarly, $(\pm 1/\sqrt{2}, 0, \pm 1/\sqrt{2})$ and $(\pm 1/\sqrt{2}, \pm 1/\sqrt{2}, 0)$ are eight more such points.

Finally, consider the instance where none of x, y, or z is zero. By the first three equations, then, $\lambda = 2x^2 = 2y^2 = 2z^2$. Putting this into the fourth equation, we obtain $3x^2=1$. This leads to eight points $(\pm 1/\sqrt{3}, \pm 1/\sqrt{3}, \pm 1/\sqrt{3})$ at which f may have its minimum value.

Now we simply check that

$$f(\pm 1,0,0) = f(0,\pm 1,0) = f(0,0,\pm 1) = 1$$

$$f(0, \pm 1/\sqrt{2}, \pm 1/\sqrt{2}) = f(\pm 1/\sqrt{2}, 0, \pm 1/\sqrt{2}) = f(\pm 1/\sqrt{2}, \pm 1/\sqrt{2}, 0) = \frac{1}{2}$$

$$f(\pm 1/\sqrt{3}, \pm 1/\sqrt{3}, \pm 1/\sqrt{3}) = \frac{1}{3}$$

to see that the minimum value of f on the sphere occurs at the eight points $(\pm 1/\sqrt{3}, \pm 1/\sqrt{3}, \pm 1/\sqrt{3})$.

The following question naturally arises: Can we determine the behavior of a function f of three variables at a critical point of the constrained optimization problem

$$\max/\min f(x,y,z)$$
$$g(x,y,z)=0 \quad ?$$

That is, do we have a second derivative test in this instance? The answer is affirmative; indeed the result is the same as for the two-variable case.

Thm. Let $f: U \subset \mathbb{R}^3 \to \mathbb{R}$ and $g: U \subset \mathbb{R}^3 \to \mathbb{R}$ be differentiable, and let $a \in U$ and $\lambda_0 \in \mathbb{R}$ be such that

$$\nabla f(a) = \lambda_0 \nabla g(a)$$
$$g(a)=0$$

If $W(a,\lambda_0) = Hf(a) - \lambda Hg(a)$ is positive definite, then f has a local minimum at a along the surface $g(x,y,z)=0$. If $W(a,\lambda_0)$ is negative definite, then f has a local maximum at a along the surface $g(x,y,z)=0$.

Proof: Let $(x,y,z) = C(s,t)$ be a parametrization of the surface $g(x,y,z)=0$ such that $C(s_a,t_a) = \vec{a}$ and suppose $\frac{\partial C}{\partial s} \times \frac{\partial C}{\partial t} \neq \vec{0}$. Define $p(s,t) = f(C(s,t))$. Then $\nabla p(s_a,t_a) = \vec{0}$. Also, it can be shown that the Hessian matrix for p at (s_a,t_a) is

$$Hp(s_a,t_a) = JC(s_a,t_a)\left(Hf(\vec{a}) - \lambda_0 Hg(\vec{a})\right)\left(JC(s_a,t_a)\right)^T$$

It can be shown that this matrix is positive- or negative definite if and only if $Hf(a) - \lambda_0 Hg(a)$ is positive- or negative definite.

In fact, this result is still true if we have a function f of n variables and a single constraint $g(x_1, x_2, \dots, x_n) = 0$ in the n variables.

Ex. In information theory, the entropy of an information source (e.g. the stream of consecutive letters in a novel) is a measure of the disorder or randomness of the source. It is defined to be

$$E(x_1, x_2, \dots, x_n) = -\sum_{i=1}^{n} x_i \ln x_i$$ where x_i is the

probability of the occurrence of the i^{th} letter in the information source. The question naturally arises: what choice of the probabilities x_i will maximize the entropy? We answer this as a constrained optimization problem:

$$\text{maximize } E(x_1, \dots, x_n)$$

$$\text{where } x_1 + x_2 + \dots + x_n = 1$$

$$\text{and } 0 < x_i < 1, \ i = 1, \dots, n.$$

First we find

$$\nabla E = -(\ln x_1 + 1, \ln x_2 + 1, \dots, \ln x_n + 1)$$

$$\nabla(x_1 + x_2 + \dots + x_n - 1) = (1, 1, \dots, 1)$$

So the possible extrema occur when $\nabla E = \lambda(1, 1, \dots, 1)$

These are solutions to

$$-\ln x_1 - 1 = \lambda$$
$$-\ln x_2 - 1 = \lambda$$
$$\vdots$$
$$-\ln x_n - 1 = \lambda$$

$$x_1 + x_2 + \ldots + x_n = 1$$

Evidently, $x_1 = x_2 = \ldots = x_n = e^{-(1+\lambda)}$. Putting this into the last equation gives

$$ne^{-(1+\lambda)} = 1$$
$$1+\lambda = -\ln(\tfrac{1}{n})$$
$$\lambda = -1 - \ln(\tfrac{1}{n})$$

and $x_1 = x_2 = \ldots = x_n = \frac{1}{n}$. Does this correspond to a maximum of E? The Hessian of E is

$$\begin{pmatrix} -1/x_1 & 0 & \cdots & 0 \\ 0 & -1/x_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & & -1/x_n \end{pmatrix}$$

and the Hessian of the constraint function is 0. Therefore, since

$$HE(\tfrac{1}{n}, \tfrac{1}{n}, \ldots, \tfrac{1}{n}) = \begin{pmatrix} -n & 0 & \cdots & 0 \\ 0 & -n & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -n \end{pmatrix}$$

is negative definite, the given values of x_i maximize the entropy. Thus the maximum entropy is

$$E\left(\tfrac{1}{n}, \ldots, \tfrac{1}{n}\right) = -\sum_{i=1}^{n} \tfrac{1}{n} \ln\left(\tfrac{1}{n}\right) = \ln n.$$

For instance, in a 26-letter alphabet, the maximum entropy that any source could produce is $\ln 26 \approx 3.258$

Three variables and two constraints

Consider the problem of finding relative extrema of $f(x,y,z)$ subject to two constraints $g_1(x,y,z)=0$ and $g_2(x,y,z)=0$. Geometrically, the two conditions mean that (x,y,z) must lie along the curve in which the two constraint surfaces intersect.

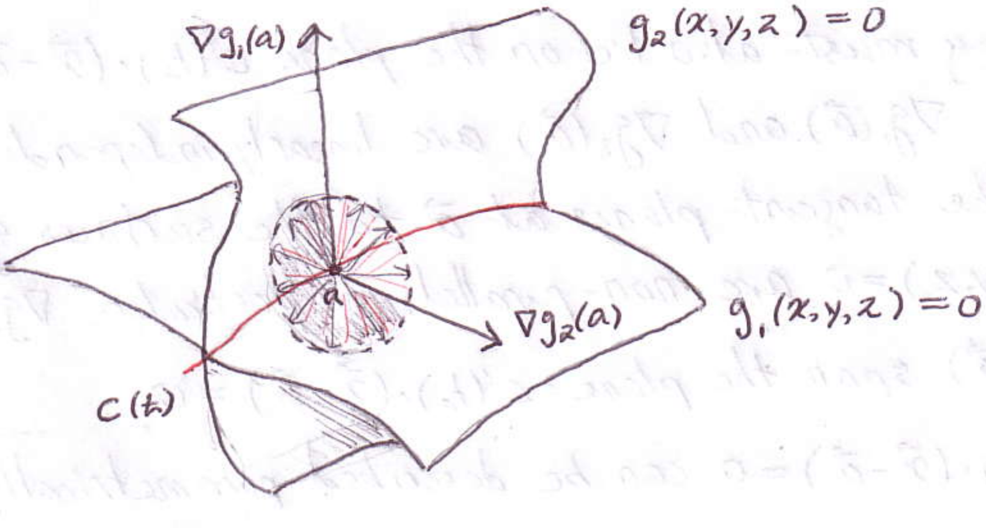


Fig. 6

Let $\vec{a}$ be a point on the constraint curve at which f has a local extremum. If $(x,y,z) = c(t)$ is a

parametrization of the curve such that $c(t_a) = \vec{a}$, then the function $p(t) = f(c(t))$ has a local extremum at t_a. Assuming all the functions involved are differentiable, we see that $0 = p'(t_a) = \nabla f(c(t_a)) \cdot c'(t_a) = \nabla f(\vec{a}) \cdot c'(t_a)$.
It follows that $\nabla f(\vec{a})$ is perpendicular to the vector $c'(t_a)$. Therefore $\nabla f(\vec{a})$ must lie on the plane $c'(t_a) \cdot (\vec{x} - \vec{a}) = 0$ (represented by the circular disk in Fig. 6).
Observe that $g_1(c(t)) = g_2(c(t)) = 0$ (why?). Thus if $c'(t_a) \neq 0$

$$0 = \nabla g_1(c(t_a)) \cdot c'(t_a) = \nabla g_1(\vec{a}) \cdot c'(t_a)$$

$$0 = \nabla g_2(c(t_a)) \cdot c'(t_a) = \nabla g_2(\vec{a}) \cdot c'(t_a)$$

Since $\nabla g_1(\vec{a})$ and $\nabla g_2(\vec{a})$ are orthogonal to $c'(t_a)$, like $\nabla f(\vec{a})$, they must also lie on the plane $c'(t_a) \cdot (\vec{x} - \vec{a}) = 0$. Moreover, $\nabla g_1(\vec{a})$ and $\nabla g_2(\vec{a})$ are linearly independent, because the tangent planes at $\vec{a}$ to the surfaces $g_1(x,y,z) = 0$ and $g_2(x,y,z) = 0$ are non-parallel. In particular, $\nabla g_1(\vec{a})$ and $\nabla g_2(\vec{a})$ span the plane $c'(t_a) \cdot (\vec{x} - \vec{a}) = 0$
(i.e. $c'(t_a) \cdot (\vec{x} - \vec{a}) = 0$ can be described parametrically as $\vec{a} + s \nabla g_1(\vec{a}) + t \nabla g_2(\vec{a})$, $s, t \in \mathbb{R}$). Since $\nabla f(\vec{a})$ is on that plane, it follows that there exist scalars $\lambda, \mu \in \mathbb{R}$ such that $\lambda \nabla g_1(\vec{a}) + \mu \nabla g_2(\vec{a}) = \nabla f(\vec{a})$
This outlines a proof of the following theorem.

Thm. Let f, g_1, and g_2 be differentiable functions on $\mathcal{U}$ in $\mathbb{R}^3$. The points in $\mathcal{U}$ satisfying the constraints

$$g_1(x,y,z)=0$$

$$g_2(x,y,z)=0$$

at which f attains local extreme values are among the solutions of the system of equations

$$\nabla f(x,y,z) = \lambda \nabla g_1(x,y,z) + \mu \nabla g_2(x,y,z)$$

$$g_1(x,y,z)=0$$

$$g_2(x,y,z)=0$$

Moreover, if $(x,y,z)=\vec{a}$, $\lambda=\lambda_0$, and $\mu=\mu_0$ is a solution to the above equations and $W(\vec{a},\lambda_0,\mu_0) = Hf(\vec{a}) - \lambda_0 Hg_1(\vec{a}) - \mu_0 Hg_2(\vec{a})$ is positive definite, then f attains a local minimum value; if $W(\vec{a},\lambda_0,\mu_0)$ is negative definite, then f attains a local maximum value.

Ex. Find the minimum value of $f(x,y,z) = x^2+y^2+z^2$ along the line at which the planes $x+y+z=1$ and $4x+2y-z=2$ intersect.

Solution: Our constraints are $g_1(x,y,z) = x+y+z-1=0$ and $g_2(x,y,z) = 4x+2y-z-2=0$. We have $\nabla f(x,y,z) = (2x, 2y, 2z)$, $\nabla g_1(x,y,z) = (1,1,1)$, and $\nabla g_2(x,y,z) = (4,2,-1)$ and we must solve

$$\nabla f = \lambda \nabla g_1 + \mu \nabla g_2$$
$$g_1 = 0, \; g_2 = 0$$

This gives the system of equations:

$$2x = \lambda + 4\mu \qquad (1)$$
$$2y = \lambda + 2\mu \qquad (2)$$
$$2z = \lambda - \mu \qquad (3)$$
$$x + y + z = 1 \qquad (4)$$
$$4x + 2y - z = 2 \qquad (5)$$

Solving the first three equations for x, y, and z, and substituting into the last two, we obtain a system of two equations in two unknowns:

$$3\lambda + 5\mu = 2$$
$$5\lambda + 21\mu = 4$$

This has solutions $\lambda = \frac{11}{19}$, $\mu = \frac{1}{19}$. Substituting these back into (1)–(3), we obtain

$$x = \frac{1}{2}(\lambda + 4\mu) = \frac{15}{38}$$
$$y = \frac{1}{2}(\lambda + 2\mu) = \frac{13}{38}$$
$$z = \frac{1}{2}(\lambda - \mu) = \frac{5}{19}$$

Since $H_{g_1} = H_{g_2} = 0$, it is easy to check in this instance that, with $(x, y, z) = \left(\frac{15}{38}, \frac{13}{38}, \frac{5}{19}\right)$, $\lambda = \frac{11}{19}$, and $\mu = \frac{1}{19}$, the matrix $W(\vec{a}, \lambda, \mu)$ is equal to

$$\begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

This matrix is positive definite, so $f\left(\frac{15}{38}, \frac{13}{38}, \frac{5}{19}\right) = \frac{247}{722}$ is a local minimum.

* Pages 25-31 were copied from "Vector Analysis" by Thomas H. Barr with only a few minor changes. The presentation that comes before is my own.

-Arkady.